Supervised Dimensionality Reduction of Hyperspectral Imagery Via Local and Global Sparse Representation

Faxian Cao , *Student Member, IEEE*, Zhijing Yang , *Member, IEEE*, Xiaobin Hong, Yongqiang Cheng , Yuezhen Huang, and Jujian Lv

Abstract—Despite the successful applications of unsupervised sparse dimensionality reduction (USDR) in pattern recognition, the USDR still suffers from two challenges for hyperspectral images (HSIs), which limit its discriminative performance: first, it cannot be applied for dimensionality reduction using both training samples and testing samples; second, it lacks the ability to integrate the spectral with spatial information for improving the discriminative performance of HSIs. In order to tackle the first challenge, we extend it to a supervised scenario, which can be applied for both training samples and testing samples, namely dimensionality reduction sparse representation (DRSR). Then, we propose a novel method called local and global DRSR (LGDRSR) to integrate the spectral information and spatial information of HSIs to further improve the discriminative performance of HSIs. The proposed LGDRSR computes the distance information between pixels of HSIs including the whole samples and their corresponding locations with a unified metric matrix. Experimental results show the proposed LGDRSR outperforms other state-of-the-art algorithms significantly.

Index Terms—Hyperspectral images (HSIs), local and global dimensionality reduction sparse representation (LGDRSR), spectral and spatial information, unsupervised sparse dimensionality reduction (USDR).

Faxian Cao is with the School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China, and also with the School of Electronic Engineering, and Computer Science, Queen Mary, University of London, E1 4NS London, U.K. (e-mail: faxiancao@foxmail.com).

Zhijing Yang is with the School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China (e-mail: yzhj@gdut.edu.cn).

Xiaobin Hong is with the School of Mechanical, and Automotive Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: mexbhong@scut.edu.cn).

Yongqiang Cheng is with the Department of Computer Science, and Technology, University of Hull, HU6 7RX Hull, U.K. (e-mail: y.cheng@hull.ac.uk).

Yuezhen Huang is with the GRGBanking Equipment Company, Ltd., Guangzhou 510663, China (e-mail: huangyz@grg.net.cn).

Jujian Lv is with the School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou 510665, China (e-mail: jujianlv@gpnu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3069030

I. INTRODUCTION

UTILIZING hyperspectral images (HSIs) technology for imaging is gaining more and more attention since the development of hyperspectral sensor technology [1], such as land cover investigation and target detection, usually with hundreds of spectral bands. The high dimensionality with hundreds of bands imposes several challenges to image classification, such as Hughes phenomenon (the problem between huge spectral bands and the limited training samples) [2] and high demands of computational resources [3].

In order to address the Hughes phenomenon [2], many algorithms have been proposed, such as sparse multinomial logistic regression (SMLR) [4], support vector machine (SVM) [5], sparse representation classification (SRC) [6], extreme learning machine [7], and collaborative representation classification [8]. Among these methods, SRC has attracted the most attentions due to its merits, i.e., sparsity and robustness [8]. The SRC represents a test sample as a linear combination of all training samples and then adopts the representation coefficients as a feature vector of the test sample [8], [9]. Finally, SRC directly assigns the class label to the test sample via calculating the minimum representation error of classes [8]. Sparse representation (SR) based methods have been applied for many applications, such as face recognition [10], brain signal processing [11], and image super-resolution [12]. Even in HSIs processing, as examples, Chen *et al.* [13] proposed a dictionary-based SR method for HSIs classification and achieved a good performance. In [14], Zhao *et al.* combined the SR and low-rank representation for HSIs denoising. In [15], Chen *et al.* proposed an SR-based method for HSIs target detection.

Although SR-based methods have achieved great success, it still faces several challenges in HSIs classification [16]. First, the huge number of spectral bands, which implies huge computation load. Second, the spatial variability of the spectral pixel, in other words, within the same categories (classes), they may have different spectral characteristic whilst the similar features may have been shared by different classes [17]. Third, the high cost of obtaining labels of training samples. Hence, it's important to reduce the dimensionality of HSIs without sacrificing significant information under the situation with limited training samples [18].

The goal of the dimensionality reduction (DR) technique is to reduce the complexity of the data [16] and preserve the desirable intrinsic information of the data simultaneously. The existing DR methods can be broadly divided into two categories, supervised methods and unsupervised ones according to whether the supervision information (class labels) is needed [18]. The examples of unsupervised dimensionality reduction (UDR) methods widely used for DR of HSIs are locality preserving projection (LPP) [19], neighborhood preserving embedding (NPE) [20], and principal component analysis (PCA) [21]. Many supervised dimensionality reduction (SDR) methods have also been proposed for DR, such as linear discriminant analysis [22], nonparametric weighted feature extraction (NWFE) [23], local fisher discriminant analysis (LFDA) [24], supervised Laplacian eigenmaps [25], and local discriminative embedding [26]. In addition, there are also some semisupervised dimensionality reduction methods being proposed for HSIs DR and achieved a good performance, such as semisupervised dimensionality reduction based on sparsity preserving [18] and semisupervised discriminant analysis (SDA) [27].

Although the methods discussed earlier have obtained good results in some applications, they still exhibit some drawbacks when being applied for HSIs classification. First, most DR-based methods consist of two steps for HSIs classification, i.e., 1) extraction of the most discriminative features from data, and 2) feeding the features to specific classifiers. Such separation often leads to limit the overall classification performance of HSIs. Second, most of the existing DR methods lack the ability to integrate the spectral and spatial information, despite that some of them use the spectral information whilst others use the spatial information for HSI classifications [28]. This will also restrict the overall classification performance of HSIs.

In [29], Zhang *et al.* proposed a UDR method based on SRC and produced a good performance, namely unsupervised sparse dimensionality reduction (USDR). However, it can only be used for unsupervised learning and, thus, some information that already existed in training samples and testing samples can'not be used for HSIs classification. On the other hand, recall the first drawback in most of the existing DR-based methods, they input the features to a specific classifier after extracting the discriminative features. This operation will limit the recognition performance. In view of these two aspects, we, hence, extend the UDR to SDR, which can be used for DR for both utilizing the information of training samples and classifying testing samples, namely dimensionality reduction based sparse representation (DRSR). Also, the proposed DRSR can solve the drawback existed in most DR methods since the proposed DRSR can be used for DR and classification simultaneously.

Furthermore, we proposed a method based on DRSR to incorporate the local and global information of HSIs namely local and global DRSR (LGDRSR) to further improve the discriminative performance of HSIs. Our LGDRSR approach has the ability to integrate the spectral information with spatial one to address the second drawback in the existing DR methods, as mentioned earlier.

The main contributions of our proposed LGDRSR can be summarized as follows.
1) Both the training samples and testing samples information have been introduced for DR and classification of HSIs.
2) The local and global spectral and spatial features in HSIs have been incorporated into the sparse optimization of HSIs for improving the performance of the DR and classification.

The rest of this work can be summarized as follows. The SRC framework is briefly reviewed in Section II. In Section III, the proposed LGDRSR will be introduced in detail. The extensive experimental results and analysis are given in Section IV. Section V concludes this article and remarks some future work.

II. RELATED WORK

A. Sparse Representation Classification

SRC aims to represent signals using as few atoms as possible in a given super complete dictionary to obtain a more concise representation of the signals, so that the features extraction becomes easier and more efficient. Thus, given a hyperspective image with N training samples from K classes of images $X = [X_1, X_2, \ldots, X_N] = [X_{N_1}, X_{N_2}, \ldots, X_{N_K}] \in R^{d \times N}$ where $N_1 + N_2 + \cdots + N_K = N$. N_K denotes the number of training samples of kth category. d is the number of the spectral bands, and K is the number of the categories. Then, for a testing sample $y \in R^{d \times 1}$, the optimization problem of SRC can be expressed as follows:

$$\hat{s} = \arg\min_s \left\{ \frac{1}{2} \|y - Xs\|_F^2 + \lambda \|s\|_1 \right\} \qquad (1)$$

where λ is a parameter that weights the importance of the two terms, s is an expansion coefficient that represents the testing sample with the training samples and F denotes the F-norm. As can be seen from (1), the first term denotes the redundancy between the testing sample y and the whole training samples X multiplies by the coefficient s. The second term denotes the regularization term, which prevents the overfitting.

In the SRC, it is assumed that the samples belonging to the same class approximately lie in a low dimensional subspace [6], and this subspace can be captured by l_1 regularization [8]. Hence, the class label la_y of testing sample y can be acquired by the following equation:

$$la_y = \arg\min_{k=1,\ldots,K} (\| y - X_{N_k} \hat{s}_{N_k} \|_2^2). \qquad (2)$$

B. Unsupervised Sparse Dimensionality Reduction

Based on SRC, USDR finds a way to reduce the dimension of HSIs with training samples. This method can be regarded as the combination of SRC and PCA. That is to say, in addition to find the spare coefficient, the USDR tries to find an additional coefficient that can be used for reducing the dimensions of HSIs. Thus, let $X = [X_1, X_2, \ldots, X_N] \in R^{d \times N}$ be N samples, and $D_i = [X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_N] \in R^d \times (N-1)$ be the collection of samples excluding the ith sample, the objective
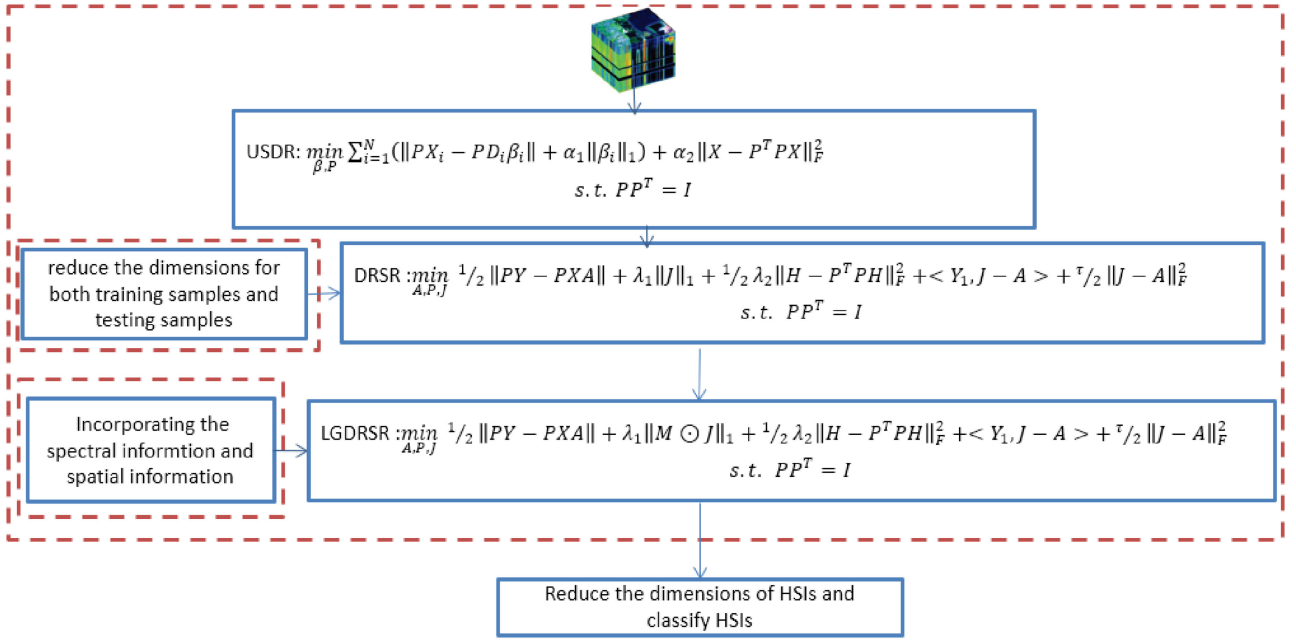
Fig. 1.   Flowchart of proposed LGDRSR.

function of USDR can be defined as follows:

$$\min_{\beta,p} \sum_{i=1}^{N} (\|PX_i - PD_i\beta_i\|_2^2 + \alpha_1\|\beta_i\|_1) + \alpha_2\|X - P^TPX\|_F^2$$

$$\text{s.t.} \, PP^T = I \tag{3}$$

where β_i is the SR coefficient of X_i over D_i, α_1 and α_2 are scalar parameters, $P \in R^{\dim \times d}$ (dim denotes the final number of dimensions) is a matrix that is used for reducing the dimensions of training samples [29], T is the matrix transpose operation. The first term $\sum_{i=1}^{N}\|PX_i - PD_i\beta_i\|_2^2$ and second term $\sum_{i=1}^{N}\alpha_1\|\beta_i\|_1$ in cost function (3) are approximation and sparse constraints, respectively. In more details, the first term denotes the redundancy between the reconstructed sample (PX_i) and the other reconstructed sample (PD_i) multiplies by the classification coefficient (β_i). The third term $\alpha_2\|X - P^TPX\|_F^2$ in cost function (3) is to ensure samples to be well reconstructed from the projected subspace by P (DR term).

The solution of (3) can be obtained by the following steps.

Step 1: Apply PCA to X in order to obtain the initial P.

Step 2: Fix P, compute each β_i by solving the following objective function:

$$\min_{\beta_i}\|PX_i - PD_i\beta_i\|_2^2 + \alpha_1\|\beta_i\|_1 \tag{4}$$

which can be solved by some convex optimization techniques or the method in [30].

Step 3: Fix X_i, update P, then the objective function can be reduced to

$$\min_{\beta,p} \sum_{i=1}^{N}\|PX_i - PD_i\beta_i\|_2^2 + \alpha_2\|X - P^TPX\|_F^2$$

$$\text{s.t.} \, PP^T = I. \tag{5}$$

This optimization problem can be solved by singular value decomposition [29].

Step 4: Go back to step 2 until the maximum number of iterations is hit or the abovementioned objective function converges.

III. Proposed LGDRSR

This section is divided into the following two subsections. First, we present the proposed DRSR, which extends the UDR [29] to SDR for simultaneous HSIs DR and classification in Section III-A. Then, based on DRSR, the proposed LGDRSR will be elaborated in Section III-B. The spectral and spatial information of HSIs will be incorporated into the proposed LGDRSR to improve the performance of DR and classification of HSIs. The corresponding flowchart of proposed LGDRSR framework can be seen in Fig. 1.

A. Dimensionality Reduction Sparse Representation

In this section, we will design the SR-based method to reduce the dimensions and classify the samples simultaneously. The USDR only uses the training samples for DR, which omits some prior information and causes the relative low recognition rate. To address this drawback, we incorporate the testing samples with the training samples to improve the efficacy of DR. Given a 2-D image $H = [X; Y] \in R^{d \times (N+n)}$ being transformed from a hyperspectral image, where there are N number of samples in the training set X and; n number of testing samples in the testing set Y, respectively. Then, the proposed DRSR can be formulated by the following equation:

$$\min_{A,P}\left\{\frac{1}{2}\|PY - PXA\|_F^2 + \lambda_1\|A\|_1 + \frac{1}{2}\lambda_2\|H - P^TPH\|_F^2\right\} \tag{6}$$

where $P \in R^{m \times d}$ is a matrix that is used for reducing the dimensions of HSIs, i.e., training samples and testing samples, λ_1 and λ_2 are two parameters, which weight the three items, A is the representation coefficient for the discriminative features acquired by the matrix P. The redundancies between reconstructed testing samples (PY) and reconstructed training samples (PX) multiply by the classification coefficient can be denoted by the first term. For the third term, we can see that the HSIs dataset H can be well reconstructed from the projected subspace by P. Furthermore, due to the Hughes phenomenon (curse of dimensionality), overfitting of classifier may occur, so the regularized term $\|A\|_1$ can help to alleviate this problem. Then, we impose the orthogonal constraint to the matrix P in order to avoid the trivial solution for the optimization problem of (6). Thus, the optimization problem can be rewritten as

$$\min_{A,P} \left\{ \frac{1}{2}\|PY - PXA\|_F^2 + \lambda_1\|A\|_1 + \frac{1}{2}\lambda_2\|H - P^TPH\|_F^2 \right\}$$

$$\text{s.t.} PP^T = I \tag{7}$$

where $I \in R^{m \times m}$ is the identity matrix. And compared with (5), which just use the training samples for DR term $\|X - P^TPX\|$, we can see that the (7) has introduced both the testing samples and testing samples information for the DR term $\|H - P^TPH\|_F^2$.

In order to solve the optimization problem of (7), we adopt the inexact augmented Lagrange multiplier (IALM) [31]–[33]. IALM is a computational framework for solving the optimization problems, especially for the statistical learning problems. The idea of IALM is to use the decomposition-coordination procedure to decompose a large global problem into multiple smaller and easier subproblems. Thus, by coordinating the solution of those subproblems, the final solution of the large global problem can be acquired. First, we introduce an auxiliary variable J to split the variable to let problem (4) become more easily solvable. Then, the equivalent model can be rewritten as follows:

$$\min_{A,P,J} \left\{ \frac{1}{2}\|PY - PXA\|_F^2 + \lambda_1\|J\|_1 + \frac{1}{2}\lambda_2\|H - P^TPH\|_F^2 \right\}$$

$$\text{s.t.} PP^T = I; J = A. \tag{8}$$

Then, the corresponding augmented Lagrangian function of model (8) can be expressed as

$$\min_{A,P,J} \left\{ \frac{1}{2}\|PY - PXA\|_F^2 + \lambda_1\|J\|_1 + \frac{1}{2}\lambda_2\|H - P^TPH\|_F^2 \right.$$

$$\left. + \langle Y_1, J - A \rangle + \frac{\tau}{2}\|J - A\|_F^2 \right\} \text{s.t.} PP^T = I \tag{9}$$

where Y_1 is the Lagrange multipliers. Then, model (9) can be solved by the alternative optimization algorithm [31]. The details are given as follows.

Update P: Fix A and J, then P can be updated as

$$P^t = \arg\min_{P} \left\{ \frac{1}{2}\|PY - PXA^t\|_F^2 + \frac{1}{2}\lambda_2\|H - P^TPH\|_F^2 \right\}$$

$$\text{s.t.} PP^T = I. \tag{10}$$

The derivation of (10) can be solved as follows:

$$P^t = \arg\min_{P} \left\{ \text{tr}(\frac{1}{2}P(Y - XA^t)(Y - XA^t)^T P^T \right.$$

$$+ \frac{1}{2}\lambda_2(H - P^TPH)(H - P^TPH)^T \Big\}$$

$$= \arg\min_{P} \left\{ \text{tr}\left(\frac{1}{2}P(Y - XA^t)(Y - XA^t)^T \right.\right.$$

$$+ \frac{1}{2}\lambda_2(HH^T - P^TPHH^T - HH^TP^TP$$

$$\left.\left. + P^TPHH^TP^TP) \right) \right\} \tag{11}$$

where tr is the trace operation [31] and t is the tth iteration. According to the characters of trace and $PP^T = I$, (11) can be rewritten as

$$P^t = \arg\min_{P} \left\{ \text{tr}\left(\frac{1}{2}P(Y - XA^t)(Y - XA^t)^T P^T \right.\right.$$

$$\left.\left. + \frac{1}{2}\lambda_2(HH^T - PHH^TP^T) \right) \right\}. \tag{12}$$

Since the HH^T has no effect to the solution of P, then (12) can be solved by

$$P^t = \underset{P}{\arg\min} \left\{ \text{tr}\left(\frac{1}{2}P(Y - XA^t)(Y - XA^t)^T \right.\right.$$

$$\left.\left. - \lambda_2 HH^T)P^T \right) \right\}. \tag{13}$$

Finally, P^t can be determined by singular value decomposition [29] to $(Y - XA^t)(Y - XA^t)^T - \lambda_2 HH^T$, which is composed by the corresponding eigenvectors of the first several smallest eigenvalues of $(Y - XA^t)(Y - XA^t)^T - \lambda_2 HH^T$. After the P^t has been acquired, which can be used for reducing the dimensions of HSIs, the variables J and A for classification of HSIs can be obtained as follows.

Update J: Fix A and P, then J can be updated as

$$J^{t+1} = \arg\min_{J} \frac{\lambda_1}{\tau^t}\|J^t\|_1 + \frac{1}{2}\|J^t - A^t + \frac{Y_1^t}{\tau^t}\|_F^2. \tag{14}$$

The solution of (14) can be solved by the soft-threshold [35] rule

$$J^{t+1} = \text{soft}\left(A^t - \frac{Y_1^t}{\tau^t} \right) = \max\left\{ 0, \text{abs}(e) - \frac{\lambda}{\tau^t} \times \text{sign}(e) \right\} \tag{15}$$

where $\text{abs}(e)$ is a function to get the absolute value of each elements in e, $e = A^t - \frac{Y_1^t}{\tau^t}$ and sign is the sign function [35].

Update A: Fix P and J, then the A can be updated as

$$A^{t+1} = \arg\min_{A}$$

$$\times \left\{ \frac{1}{2}\|P^tY - P^tXA\|_F^2 + \frac{\tau^t}{2}\|J^t - A^t + \frac{Y_1^t}{\tau^t}\|_F^2 \right\}. \tag{16}$$

The solution of (16) can be acquired by the first-order derivation

$$A^{t+1} = (X^T P^{t^T} P^t X + \tau^t I)^{-1} (X^T P^{t^T} P^t Y + \tau^t J^t + Y_1^t). \quad (17)$$

B. Local and Global DRSR

In the abovementioned Section III-A, we entailed the DRSR method to reduce the dimensions of HSIs and classify the HSIs simultaneous. Inspired by Wang *et al.* [31], which introduced a new distance metric to low-rank representation for HSIs classification, we adopt this distance metric in our proposed DRSR, which can further preserve the intrinsic information and integrate the spectral and spatial information when reducing the dimensions of HSIs.

Given a hyperspectral image $H = [X; Y] = [X_1, X_2, \ldots, X_N; Y_1, \ldots, Y_n] \in R^{d \times (N+n)}$ where X_i and Y_j $(i = 1, 2, \ldots, N; \ j = 1, 2, \ldots, n)$ are the spectral column vector of training samples and testing samples respectively. Denote the spatial feature matrix by $L = [L_1, L_2, \ldots, L_N, L_{N+1}, \ldots, L_{N+n}] \in R^{2(N+n)}$ where L_i $(i = 1, 2, \ldots, N + n)$ is the position coordinate of the ith spectral pixel. It should be noted that the values of X_i, Y_i, and L_i are normalized to the range of [0,1]. Then, a good way to integrate the spectral and spatial information of HSIs is to compute the distance matrix, which can be formulated as

$$M_{i,j} = \sqrt{\|X_i - Y_j\|_2^2 + m\|L_i - L_j\|_2^2} \quad (18)$$

where $i = 1, 2, \cdot, N; \ j = 1, 2, \cdot, n$; m is a parameter to control the weight of the spectral and spatial distance. For each training sample in a hyperspectral image, the training samples located either near to or far away from the testing samples, if the training samples are near to testing samples, we can regard this type as the local information, otherwise, this can be regarded as the global information. If the spectral information ($\|X_i - X_j\|_2^2$) and spatial information ($\|L_i - L_j\|_2^2$) can be incorporated into the DRSR, the DR and classification performance can be improved. From (18), we can see that M contains the local and global information of HSIs since both the spectral distance value $\|X_i - X_j\|_2^2$ and the spatial distance value $\|L_i - L_j\|_2^2$ (it's a feature of relative Euclidian distance between the locations of pixels) contain local and global spectral and spatial information of a training and testing pixel in HSIs. By constructing the matrix M, and introducing a new variable J, the LGDRSR can be written as

$$\min_{A,P,J} \left\{ \frac{1}{2} \|PY - PXA\|_F^2 + \lambda_1 \|M \odot J\|_1 \right.$$

$$\left. + \frac{1}{2} \lambda_2 \|H - P^T PH\|_F^2 + \langle Y_1, J - A \rangle + \frac{\tau}{2} \|J - A\|_F^2 \right\}$$

$$\text{s.t. } PP^T = I \quad (19)$$

where $\odot$ is the Hadamard operator [34]. Then, the solution of LGDRSR can be obtained as follows.

Update P: Fix A and J, then the matrix P can be updated as

$$P^t = \arg \min_P \left\{ \frac{1}{2} \|PY - PXA^t\|_F^2 + \frac{1}{2} \|H - P^T PH\|_F^2 \right\}$$

$$\text{s.t. } PP^T = I. \quad (20)$$

The solution of P at the tth iteration can be acquired by computing SVD of $(Y - XA^t)(Y - XA^t)^T - \lambda_2 HH^T$. The size of P^t is composed by the dim eigenvectors of the first smallest dim eigenvalues of $(Y - XA^t)(Y - XA^t)^T - \lambda_2 HH^T$ (dim denotes the number of dimensions, such as 5, 10, and 15).

Update J: Fix A and P, then J can be updated as

$$J^{t+1} = \arg \min_J \frac{\lambda_1}{\tau^t} \|M \odot J^t\|_1 + \frac{1}{2} \|J^t - A^t + \frac{Y_1^t}{\tau^t}\|_F^2. \quad (21)$$

The solution of (21) can be solved by the transformation of soft-threshold rule

$$J^{t+} = \text{soft} \left(A^t - \frac{Y_1^t}{\tau^t}, M \times \frac{\lambda_1}{\tau^t} \right)$$

$$= \max \left\{ 0, abs(e) - \left(M \times \frac{\lambda_1}{\tau^t} \right) \right\} \times \text{sign}(e) \quad (22)$$

where $e = A^t - \frac{Y_1^t}{\tau^t}$.

Update A: Fix P and J, then A can be updated as

$$A^{t+1} = \arg \min_A$$

$$\left\{ \frac{1}{2} \|P^t Y - P^t XA\|_F^2 + \frac{\tau^t}{2} \|J^{t+1} - A + \frac{Y_1^t}{\tau^t}\|_F^2 \right\}. \quad (23)$$

The solution of (23) can be achieved by

$$A^{t+1} = (X^T P^{t^T} P^t X + \tau^t I)^{-1}$$

$$(X^T P^{t^T} P^t Y + \tau^t J^{t+1} + Y_1^t). \quad (24)$$

Finally, the overall optimization problem for solving the proposed LGDRSR is described in Algorithm 1.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

This section has been divided into the following five sections. First, two benchmarking HSIs datasets have been described in Section IV-A followed by benchmarking approaches in Section IV-B. In Section IV-C, the parameters and our contributions have been evaluated with four experiments. The effectiveness of different dimensions and training samples on our proposed LGDRSR, and the comparison with other state-of-the-art algorithms have been presented in Sections IV-D and IV-E, respectively.

A. HSIs Datasets

1) Indian Pines: The airborne visible infrared imaging spectrometer sensor captured this image in northwestern Indiana USA with 145×145 pixels and each pixel has 200 spectral bands after removing 20 noisy bands. The wavelength ranges of

---

Algorithm 1: The proposed LGDRSR.

Input: Hyperspectral image: $H = [X; Y] \in R^{d \times (N+n)}$;
N training samples from K classes of H:
$X = [X_{N_1}, X_{N_2}, \ldots, X_{N_K}] \in R^d \times N$;
n testing samples: $Y = [Y_1, Y_2, \cdots, Y_n] \in R^{d \times n}$;
spatial feature matrix:
$L = [L_1, L_2, \ldots, L_N, L_{N+1}, \ldots, L_{N+n} \in R^{2 \times (N+n)}]$;
Parameters:$\lambda_1, \lambda_2, \tau^t, m, Y_1 = 0$,
$J^t = A^t = (X^T X + 2^{-10} I)^{-1} X^T Y$.
1.1 Set $t = 0$
Update P: $P^t =$
$arg\,\underset{P}{min}\{\frac{1}{2}\|PY - PXA\|_F^2 + \lambda_2\|H - P^T PH\|_F^2\}$
$s.t.\ PP^T = I \Longrightarrow P^t \leftarrow$
$SVD((Y - XA^t)(Y - XA^t)^T - \lambda_2 HH^T)$:
the corresponding dim eigenvectors of the first dim
 smallest eigenvalues
1.3 Update J: $\quad\quad J^{t+1} = \underset{J}{min}J^{t+1} =$
$\quad arg\,\underset{J}{min}\frac{\lambda_1}{\tau^t}\|M \odot J^t\|_1 + \frac{1}{2}\|J^t - A^t + \frac{Y_1^t}{\tau^t}\|_F^2$
$\quad\quad \Longrightarrow J^{t+1} = soft(A^t - \frac{Y_1^t}{\tau^t}, M \times \frac{\lambda_1}{\tau^t})$.
1.4: Update A:
 $A^{t+1} =$
$arg\,\underset{A}{min}\|P^t Y - P^t XA\|_F^2 + \frac{\tau^t}{2}\|J^{t+1} - A + \frac{Y_1^t}{\tau^t}\|_F^2$
$\Longrightarrow A^{t+1} = (X^T P^{t^T} P^t X + \tau^t I)^{-1}(X^T P^{t^T} P^t Y +$
$\tau^t J^{t+1} + Y_1^t)$.
1.5 Update other parameters:
$Y_1^{t+1} = Y_1^t + \tau^t(J^{t+1} - A^{t+1}); \tau^{t+1} = 1.1 \times \tau^t$.
1.6 Quit the algorithm if the stopping criterion is met;
 otherwise, go back to Step 1.2.
1.7 Predict the testing sample label:
$la_{Y_i^*} = arg\,\underset{K}{min}(\|X\hat{\alpha} - X_{N_k}\hat{\alpha}_{N_k}\|_2^2); i =$
$1, \ldots, n, k = 1, \ldots, K.$

---

the spectral bands ranges from 0.4 to $2.4\,\mu m$. In addition, there are 16 classes with 10 366 pixels need to be classified [36], [37].

2) Pavia University: The reflective optics system imaging spectrometer sensor captured this image in the urban area of Pavia University Italy with 610×340 pixels and each pixel has 103 spectral bands after removing 12 noisy bands. The wavelength of the spectral band ranges from 0.43 to $0.86\,\mu m$. In addition, there are 9 classes with 42 776 pixels need to be classified [38], [39].

B. Benchmarking Approaches

We have selected state-of-the-art DR methods for comparison with the proposed LGDRSR including PCA [21], NWFE [23], linear discriminant embedding (LDE) [40], regularization LDE (RLDE) [40], LFDA [24], SDA [27], semisupervised local discriminative analysis LPP (SELDLPP) [40], and semisupervised local discriminative analysis NPE (SELDNPE) [41]. The LIBSVM [42] software is used for implementation of

the kernel SVM (KSVM). The MATLAB code of abovementioned methods can be downloaded at [43]. For the parameter setting of abovementioned methods (KSVM-PCA, KSVM-NEFE, KSVM-LDE, KSVM-RLDE, KSVM-LFDA, KSVM-SDA, KSVM-SELDLPP, and KSVM-SELDNPE), they follow on the setting of [40].

All the experiments are conducted using MATLAB R2015a running on a computer with 2.9 GHz i7 7820HQ CPU and 32 GB RAM. The training samples (up to 50% in each class) are randomly generated from HSIs and the remaining for testing. All the experimental results are repeated 10 times and the results are averaged to obtain classification accuracies. The performance has been measured in terms of overall accuracies (OA), average accuracies (AA), category accuracies (CA), kappa coefficient (k), and computational time. The computational time includes training time (Tr) and testing time (Ts).

OA is the ratio of the total number of correctly classified testing samples over the total number of the testing samples; CA is the ratio of the number of correctly classified testing sample in each class over the number of testing sample in each class; AA is the mean of accuracies achieved in each class and k is the statistics computed by weighting the measured accuracies.

C. Parameters and Contributions Analysis

In this section, we will analyze the impact of the parameters and the contributions of the proposed LGDRSR with four experiments, including λ_1, λ_2, and τ^t in (19), and m in (18). We also applied some DR methods to SRC for the performance comparison, including PCA, NWFE, and LDE. All the experimental results are carried out with five training samples per class and we set the target dimension to be 10. Besides, we will also apply the SR for comparison, which means it will use the whole dimensions in this method.

1) Experiment #1: In this experiment, we evaluate $\lambda_2 = 2^{a2}$ where $a2$ ranges in $[-15, -14, \ldots, 10]$ while fixing $\lambda_1 = 2^{-10}$, $\tau^0 = 2^{-10}$ for the proposed DRSR, $\lambda_1 = 2^2$, $\tau^0 = 2^{-10}$, and $m = 20$ for the proposed LGDRSR ($m = 0$ means that the proposed LGDRSR just use the spectral information). From Fig. 2, we can see that $a2$ has some effect on DRSR and LGDRSR ($m = 0$) in both Indian Pines and Pavia University dataset. The results become more stable while incorporating the spatial information into the LGDRSR. Hence, we set $a2$ to be 0 for DRSR, LGDRSR ($m = 0$), and LGDRSR in both Indian Pines and Pavia University.

2) Experiment #2: In this experiment, we evaluate $\lambda_1 = 2^{a1}$ where a1 ranges in $[-20, -19, \ldots, 5]$ while fixing $\lambda_1 = 2^{-10}$, $\tau^0 = 2^{-10}$ for the proposed DRSR, $\tau^0 = 2^{-10}$ for SR, SR-PCA, SR-LDE, and SR-NWFE, $\lambda_1 = 2^2$, $\tau^0 = 2^{-10}$, and $m = 20$ for the proposed LGDRSR. From Fig. 3, we can see that the SR-PCA, SR-LDE, and SR-NWFE produce lower classification accuracies than DRSR. We can also observe that the LGDRSR ($m = 0$) has achieved better accuracies than DRSR and LGDRSR ($m = 20$) has achieved higher accuracies than LGDRSR ($m = 0$) when incorporated with the spectral and spatial information. Also, we will set $a1 = -8$ for SR, SR-PCA, DRSR, and LGDRSR ($m = 0$), $a1 = 1$ for LGDRSR in Indian
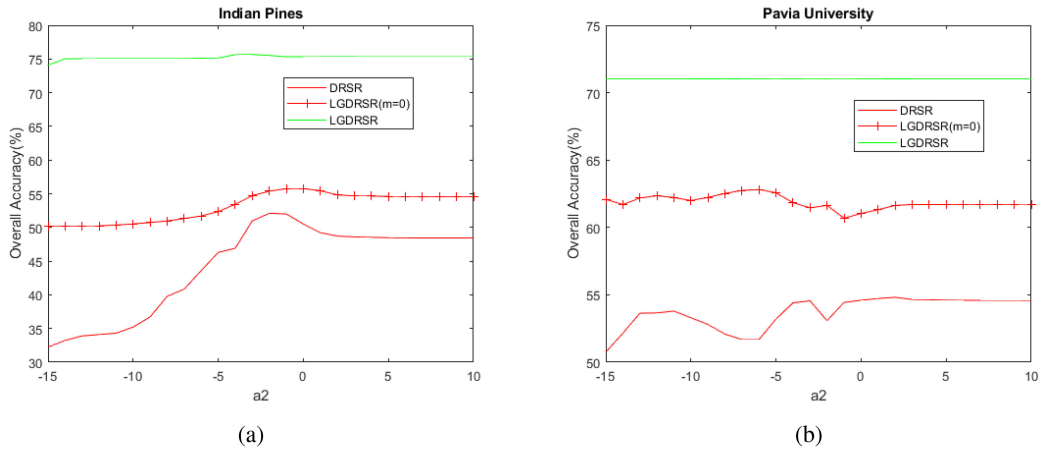
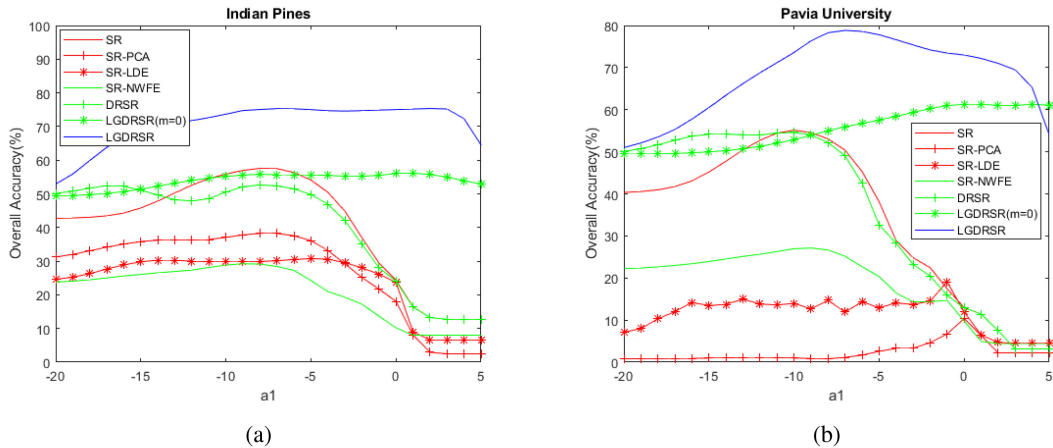Fig. 2. Effect of key parameters of $\frac{\lambda_2}{a2}$ on Indian Pines (left) and Pavia University (right).

Fig. 3. Effect of key parameters of $\frac{\lambda_1}{a1}$ on Indian Pines (left) and Pavia University (right).

Pines. In Pavia University, a1 will be set to -10 for SR and DRSR, $a1 = 0$ for SR-PCA and LGDRSR ($m = 0$), $a1 = -7$ for LGDRSR. For both Indian Pines and Pavia University dataset, $a1$ is set to -8 and -9 for SR-LDE and SR-NWFE, respectively.

3) Experiment #3: In this experiment, we evaluate the parameter $\tau = 2^{a3}$ ($a3$ belongs to $[-15, \ldots 5]$) where m is set to 20 for the proposed LGDRSR. From Fig. 4, we can see that the proposed DRSR obtains better recognition rate than others, including SR-PCA, SR-LDE, SR-NWFE, etc. It can also be seen that the LGDRSR ($m = 0$) and LGDRSR achieved higher classification accuracies than DRSR when the spectral and spatial information are incorporated for DR. In the following experiments, we will set $a3 = -4$ for SR-PCA and the proposed DRSR, $a3 = -2$ for SR-LDE, $a3 = -10$ for SR-NWFE, LGDRSR ($m = 0$), and LGDRSR ($m = 20$) in Indian Pines dataset. In Pavia University dataset, we set $a3 = 3$ for SR-PCA, $a3 = -10$ for SR-LDE, SR-NWFE, DRSR, LGDRSR ($m = 0$), and LGDRSR ($m = 20$).

4) Experiment #4: In this experiment, we will evaluate the impact of parameter m. From Fig. 5, we can see that the

proposed LGDRSR becomes stable when $m \geq 5$, which shows a good performance of the proposed LGDRSR. In the following experiment, we will set $m = 30$ in both Indian Pines dataset and Pavia University dataset.

D. Effect of Different Dimensions and Training Samples

In this section, we further evaluate the performance in dimensions reduction and classification of the proposed DRSR and LGDRSR by varying the dimensions of samples and the number of training samples. The training samples vary between 5, 10, and 15 while the dimensions vary between 10, 15, and 20 (Q and Dim denotes the number of training samples and dimensions, respectively in the table). As can be seen from Tables I and II, the classification accuracies achieved by the proposed DRSR in Indian Pines and Pavia University datasets are worse than KSVM-NWFE, but more stable and better than SR-PCA when varying the number of training samples and dimensions. The classification accuracies of the proposed DRSR are higher than SR-PCA, with more than 10% in each metric (OA, AA, and k). The proposed LGDRSR can further improve the performance
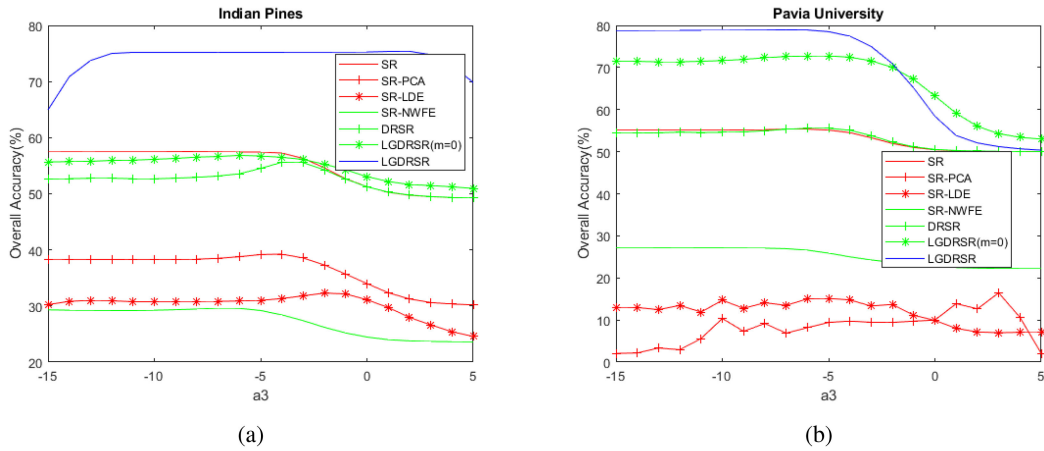
(a)　　　　　　　　　　　　　　　　　　　　　(b)

Fig. 4.　Effect of key parameters of $\frac{\tau}{a3}$ on Indian Pines (left) and Pavia University (right).

TABLE I

CLASSIFICATION ACCURACY (%) AND STANDARD DEVIATION FOR TESTING SAMPLES WITH DIFFERENT DIMENSION AND TRAINING
SAMPLES PER CLASS FOR INDIAN PINES DATASET

	Dim	10				15				20			
Q	Index	SR-PCA	KSVM-NWFE	DRSR	LGDRSR	SR-PCA	KSVM-NWFE	DRSR	LGDRSR	SR-PCA	KSVM-NWFE	DRSR	LGDRSR
5	OA	39.2±13.1	59.3±4.1	56.0±3.5	**75.4±2.0**	43.9±11.9	58.5±4.2	56.4±3.0	**75.4±2.0**	42.5±13.1	59.1±3.8	56.7±2.8	**75.4±2.0**
	AA	43.9±17.5	71.2±1.2	66.1±2.6	**83.9±1.5**	54.0±11.9	71.3±1.3	67.9±2.0	**83.9±1.5**	48.8±20.1	71.5±1.3	68.5±2.0	83.9±1.5
	k	33.2±13.8	54.4±4.5	51.3±3.4	**72.3±2.2**	38.4±12.3	53.7±4.4	51.7±3.0	**72.3±2.2**	36.7±14.2	54.3±4.1	52.1±2.7	**72.3±2.2**
10	OA	43.3±24.5	68.1±3.8	63.1±2.7	**86.7±1.5**	43.2±25.5	68.8±3.3	64.0±2.6	**86.7±1.5**	43.0±28.0	68.5±3.0	64.6±2.7	**86.7±1.5**
	AA	46.1±26.2	78.2±2.0	70.9±2.0	**91.2±0.9**	48.2±29.9	79.1±2.1	73.9±1.5	**91.2±0.9**	50.9±31.6	79.1±1.8	74.9±1.6	**91.2±0.9**
	k	37.6±25.3	64.1±4.1	58.5±2.9	**84.9±1.8**	37.5±26.6	65.0±3.6	59.5±2.9	**84.9±1.7**	37.6±29.1	64.6±3.2	60.3±3.0	**84.9±1.7**
15	OA	47.9±19.9	71.3±3.1	64.5±2.9	**89.2±1.8**	49.3±22.7	70.9±3.4	65.8±2.7	**89.2±1.8**	43.6±29.1	71.1±3.3	66.7±2.6	**89.2±1.8**
	AA	47.1±22.8	80.8±1.5	70.2±2.1	**93.3±1.1**	51.6±29.2	81.2±1.4	75.1±1.4	**93.3±1.1**	48.6±34.0	81.2±1.6	76.2±1.4	**93.3±1.1**
	k	41.6±21.0	67.8±3.4	59.9±3.0	**87.8±2.0**	43.0±24.8	67.4±3.7	61.5±2.8	**87.8±2.0**	37.7±30.9	67.6±3.5	62.5±2.7	**87.8±2.0**

TABLE II

CLASSIFICATION ACCURACY (%) AND STANDARD DEVIATION FOR TESTING SAMPLES WITH DIFFERENT DIMENSION AND TRAINING
SAMPLES PER CLASS FOR PAVIA UNIVERSITY DATASET

	Dim	10				15				20			
Q	Index	SR-PCA	KSVM-NWFE	DRSR	LGDRSR	SR-PCA	KSVM-NWFE	DRSR	LGDRSR	SR-PCA	KSVM-NWFE	DRSR	LGDRSR
5	OA	16.4±8.6	63.1±7.8	55.5±4.6	**78.7±3.5**	16.4±8.5	62.9±7.4	55.1±4.7	**78.7±3.4**	16.8±8.9	63.2±7.5	55.1±4.6	**78.7±3.5**
	AA	14.7±5.5	74.7±3.2	56.2±3.8	**82.8±2.1**	14.6±5.5	74.8±3.2	56.5±3.7	**82.6±2.1**	14.7±5.5	74.7±3.2	56.6±3.5	**82.6±2.1**
	k	4.5±5.8	54.7±8.2	44.5±4.5	**73.0±4.0**	4.4±5.7	54.6±7.8	44.3±4.5	**73.0±4.0**	4.7±6.0	54.8±7.9	44.2±4.3	**73.0±4.0**
10	OA	17.6±12.6	69.0±3.4	60.2±3.8	**83.7±4.6**	17.7±12.7	70.9±2.5	60.9±3.8	**83.6±4.7**	17.7±12.8	70.1±1.9	61.2±3.7	**83.6±4.7**
	AA	14.8±4.2	79.5±1.0	57.0±3.5	**87.7±2.3**	14.8±4.2	80.3±1.6	58.3±2.9	**87.5±2.2**	14.8±4.2	79.9±1.3	58.6±3.0	**87.5±2.2**
	k	3.4±8.3	61.5±3.3	48.9±3.9	**79.2±5.3**	3.5±8.5	63.6±2.8	49.9±3.9	**79.1±5.4**	3.5±8.5	62.7±2.0	50.2±3.9	**79.1±5.4**
15	OA	17.1±9.5	74.6±4.4	63.1±5.1	**88.6±1.2**	17.2±9.6	75.7±3.9	64.9±3.7	**88.4±1.3**	17.4±9.8	75.6±3.6	65.4±3.5	**88.4±1.3**
	AA	22.9±3.0	81.6±1.8	59.2±2.7	**90.9±1.0**	22.9±3.0	82.4±1.8	61.4±2.5	**90.6±1.3**	23.0±3.0	82.3±1.8	62.0±2.3	**90.5±1.3**
	k	9.4±5.3	67.7±5.1	52.4±5.2	**85.3±1.5**	9.5±5.4	69.1±4.6	54.5±3.9	**85.0±1.6**	9.6±5.5	68.9±4.3	55.1±3.6	**85.0±1.6**

in term of OA, AA, and k, about 20% higher than DRSR. This further shows the outstanding performance of the proposed DRSR and LGDRSR.

E. Comparison With Other State-of-the-Art Algorithms

In this section, we evaluate our proposed methods by comparing them with a list of state-of-the-art methods in three scenarios: fixed training samples, fixed dimensions, and fixed both training samples and dimensions.

First, we fix the number of training samples (5 per class) to show the effect of dimensions on classification accuracies. As can be seen from Fig. 6, the proposed DRSR acquired worse accuracies than some state-of-the-art algorithms in some cases, this is because proposed DRSR is just utilizes the spectral information of HSIs. However, by incorporating the spectral-spatial information, our proposed LGDRSR (the error bar of the

TABLE III
CLASSIFICATION ACCURACY (%) AND STANDARD DEVIATION FOR TESTING SAMPLES WITH DIFFERENT TRAINING SAMPLES PER CLASS FOR INDIAN PINES DATASET

| Q	Index	DR+KSVM classifier								LGDRSR (m=0)	LGDRSR
		PCA	NWFE	LDE	RLDE	LFDA	SDA	SELDLPP	SELDNPE		
5	OA	52.3±4.7	55.5±4.1	49.0±7.6	41.7±9.2	46.1±4.7	55.4±3.9	50.9±6.0	52.7±4.9	55.2±3.0	**75.4±2.0**
	AA	64.0±2.5	67.9±1.4	61.9±6.1	51.9±10.1	59.2±3.9	67.1±2.9	62.5±3.9	64.2±4.0	66.9±1.9	**83.8±1.5**
	k	46.9±4.9	50.4±4.3	43.7±7.9	36.0±9.1	40.3±4.9	50.1±4.3	45.3±6.5	47.2±5.3	50.1±3.1	**72.3±2.2**
10	OA	60.9±4.6	63.0±4.4	56.1±5.0	47.1±3.5	41.8±2.7	61.8±5.1	57.9±4.4	60.9±3.2	61.4±2.9	**86.6±1.6**
	AA	71.4±2.8	74.2±2.6	68.3±3.4	59.7±4.4	53.5±2.8	72.6±2.5	69.1±2.4	71.2±1.9	72.2±1.2	**91.1±1.0**
	k	56.2±4.9	58.5±4.9	51.4±5.3	41.7±3.6	35.7±2.8	57.1±5.5	52.8±4.5	56.2±3.5	56.7±3.1	**84.8±1.8**
15	OA	63.7±4.0	67.5±3.8	57.9±2.9	56.5±2.6	41.8±3.0	63.4±2.6	61.6±2.2	64.6±2.4	65.1±1.8	**89.1±1.7**
	AA	73.7±2.9	78.2±2.7	70.2±3.3	67.9±2.0	53.5±2.4	75.0±1.5	71.1±1.7	73.9±2.2	74.9±1.0	**93.1±1.1**
	k	59.4±4.4	63.6±4.2	53.4±3.1	51.6±2.7	35.8±3.2	59.0±2.8	56.9±2.6	60.3±2.7	60.9±2.0	**87.6±2.0**
20	OA	66.6±2.8	69.4±2.9	60.9±3.1	59.1±2.4	45.4±3.0	66.1±3.2	64.2±2.7	65.3±3.0	66.8±2.3	**90.8±0.7**
	AA	76.5±2.2	80.3±1.2	72.7±2.9	71.0±1.8	58.3±3.9	77.9±1.8	74.6±1.5	75.3±2.8	76.8±1.2	**94.4±0.7**
	k	62.5±3.0	65.6±3.2	56.3±3.3	54.2±2.6	39.5±3.1	62.0±3.5	59.8±3.0	61.0±3.4	62.8±2.4	**89.6±0.8**
25	OA	67.7±2.1	72.2±1.6	62.0±3.3	62.1±2.5	51.7±3.3	68.9±1.9	66.9±1.8	67.6±1.6	68.5±1.7	**91.9±0.8**
	AA	77.4±1.7	82.3±1.1	74.0±3.8	73.1±1.9	64.8±3.2	79.5±1.6	76.9±1.0	77.1±1.2	78.6±1.0	**95.3±0.6**
	k	63.7±2.3	68.7±1.7	57.5±3.6	57.5±2.6	46.3±3.5	65.0±1.9	62.7±1.9	63.5±1.7	64.6±1.9	**90.8±0.9**
30	OA	70.0±1.9	73.7±0.9	61.2±5.0	61.2±2.2	54.1±1.5	70.0±1.3	67.9±1.6	67.7±1.8	69.9±1.7	**93.1±0.8**
	AA	79.3±1.9	83.1±0.8	73.5±3.2	72.5±2.3	66.8±1.6	80.9±1.3	77.7±1.2	77.6±1.2	79.4±1.3	**95.7±0.6**
	k	66.2±2.1	70.3±1.0	56.7±5.4	56.6±2.4	48.7±1.6	66.2±1.4	63.9±1.8	63.7±2.0	66.2±1.8	**92.1±0.9**
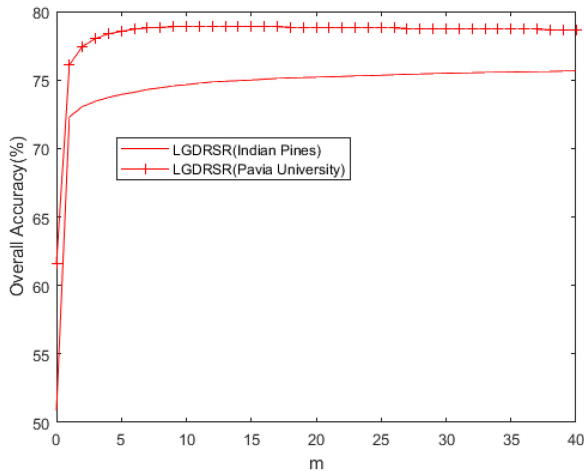
Fig. 5.　Effect of key parameters of m in Indian Pines and Pavia University.

proposed LGDRSR has been plotted in Fig. 6) has achieved better classification accuracies than the other state-of-the-art methods with different dimensions in both Indian Pines and Pavia University dataset. It is worth noting that the proposed LGDRSR can achieve very a good performance even at lower dimension (below 5). This attributes to the capability of our algorithm to incorporate the spectral and spatial information to improve the DR and the classification performance. It verifies that the proposed LGDRSR has a good performance.

Second, we fix the dimensions (5) to show the effect of the number of training samples on classification accuracies. As can be seen from Tables III and IV, the proposed LGDRSR ($m = 0$) has obtained a good performance than some state-of-the-art algorithms while the performance is worse than some other

algorithms in some experimental conditions. This indicates that the proposed LGDRSR ($m=0$) has similar DR and classification capacity with these state-of-the-art algorithms since only the spectral information has been used in proposed LGDRSR ($m = 0$). However, with integrating the spectral and spatial information simultaneously, the proposed LGDRSR achieves a significant improvement on classification accuracies and outperforms other state-of-the-art methods when varying the training samples. This further verifies the performance of the proposed LGDRSR.

Third, we fix the dimensions (5) and 1% training samples (TR) to show the classification accuracies by using the metrics, OA, AA, k, and CA, and the remaining testing samples (TS) for testing. We also report the time consumed by the algorithms for comparison. As can be seen from Tables V and VI, the situation is similar with the Tables III and IV for proposed LGDRSR ($m = 0$), and with integrating the spectral and spatial information, the proposed LGDRSR has outperformed other state-of-the-art methods in terms of recognition rates. Figs. 7 and 8 show the corresponding classification maps. It can be seen there are many classification errors existed in the classification maps of the other comparison methods while the ones obtained by the proposed LGDRSR have better clarity (less classification errors than others). This again verifies the proposed LGDRSR.

At the last three rows of Tables V and VI, we report the computation time (DR time, training time, and testing time of each method). It can be seen that the proposed LGDRSR requires more processing time than other methods in an acceptable range in both Indian Pines and Pavia University dataset. Given the much better classification accuracies achieved by our proposed method and the accessibility of the ever increasing processing power, the proposed LGDRSR presents a strong performance
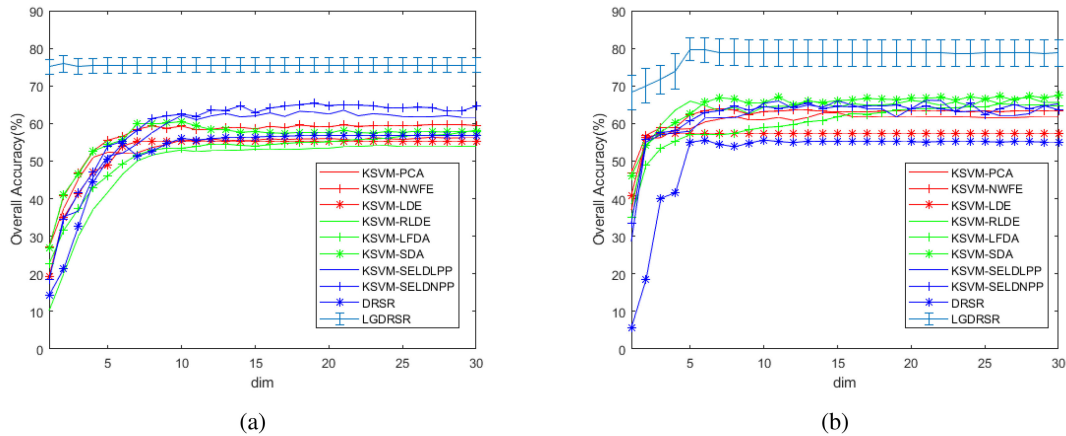
Fig. 6. Classification accuracies with five training samples per class under different dimensions. (a) Indian Pines dataset. (b) Pavia University dataset.

TABLE IV
CLASSIFICATION ACCURACY (%) AND STANDARD DEVIATION FOR TESTING SAMPLES WITH DIFFERENT TRAINING SAMPLES PER
CLASS FOR PAVIA UNIVERSITY DATASET

Q	Index	DR+KSVM classifier								LGDRSR (m=0)	LGDRSR
		PCA	NWFE	LDE	RLDE	LFDA	SDA	SELDLPP	SELDNPE		
5	OA	58.6±6.5	62.2±6.1	57.2±7.1	65.9±8.8	57.0±4.8	63.5±6.8	59.3±7.5	60.1±7.1	59.2±4.5	**79.7±3.1**
	AA	70.7±3.3	73.5±3.1	70.1±2.8	73.1±3.5	63.3±1.9	74.7±3.1	71.4±3.7	72.2±3.8	71.9±2.3	**84.2±2.0**
	k	49.3±6.6	53.4±6.4	47.8±6.8	57.5±9.2	46.8±4.6	55.2±7.1	50.2±7.5	51.1±7.1	50.2±4.8	**74.2±3.6**
10	OA	64.7±5.1	66.9±3.0	58.9±6.5	70.9±2.9	60.5±4.2	67.1±7.1	66.0±4.0	67.1±1.7	65.4±3.5	**84.0±4.8**
	AA	76.2±1.8	78.1±1.4	71.7±3.1	79.7±2.0	68.3±2.0	79.4±2.3	77.8±1.7	77.4±1.6	75.0±1.7	**88.9±3.0**
	k	56.4±5.2	59.1±3.0	50.0±6.6	63.7±3.3	51.2±4.2	59.7±7.4	58.2±4.0	59.2±1.6	56.9±3.7	**79.7±5.7**
15	OA	67.7±5.3	69.4±4.4	59.6±5.0	73.3±3.2	67.1±3.3	71.6±4.5	71.3±5.2	71.0±4.6	69.1±3.1	**88.2±1.0**
	AA	77.1±1.7	79.6±1.4	72.5±3.0	81.1±1.3	73.8±2.2	82.1±1.6	79.3±2.1	78.8±2.1	76.9±1.0	**91.9±1.1**
	k	59.6±5.7	61.9±4.6	50.8±5.1	66.3±3.4	58.6±3.8	64.9±4.8	63.9±5.6	63.5±5.1	61.1±3.4	**84.8±1.3**
20	OA	67.7±2.4	70.2±6.6	62.5±3.2	73.1±3.3	69.1±4.0	71.8±4.1	69.8±1.9	70.4±2.7	69.4±2.1	**91.5±2.0**
	AA	78.4±1.0	80.4±2.4	74.6±2.3	82.5±1.5	77.2±2.8	81.7±1.8	79.6±1.4	79.7±1.9	77.6±1.3	**93.5±0.9**
	k	59.8±2.5	63.1±6.9	53.9±3.3	66.3±3.9	61.3±4.6	65.0±4.3	62.4±2.2	63.0±2.9	61.4±2.4	**88.9±2.6**
25	OA	69.8±3.0	73.1±4.5	65.5±1.7	75.0±2.7	72.0±2.6	75.5±3.1	73.9±2.8	72.0±2.9	72.4±2.6	**92.4±1.7**
	AA	80.3±1.6	82.0±0.7	76.4±2.0	83.9±1.1	80.1±1.3	83.5±1.5	82.0±1.0	80.5±1.5	79.5±1.1	**94.6±0.7**
	k	62.4±3.2	66.4±4.6	57.3±1.5	68.7±2.9	65.0±2.8	69.2±3.5	67.1±2.9	65.0±3.0	65.0±2.9	**90.1±2.2**
30	OA	70.4±2.3	75.0±1.9	66.8±4.9	77.1±3.8	74.6±2.6	76.5±2.0	72.8±1.1	74.2±2.0	72.8±1.3	**92.6±1.6**
	AA	80.6±1.0	82.9±0.7	76.7±2.9	83.9±1.5	81.7±1.6	84.5±0.9	82.0±1.2	81.7±1.1	79.6±0.6	**94.9±0.8**
	k	63.0±2.5	68.4±2.1	58.6±4.9	70.9±4.4	67.9±3.0	70.3±2.2	65.9±1.2	67.4±2.3	65.3±1.4	**90.3±2.1**

improvement over other candidate algorithms for HSIs classifications.

F. Extended Experiments and Analysis

In this section, we have conducted extended comparison experiments to further validate the performance of the proposed DRSR and LGDRSR.

We add noise (independent identically distributed: zero mean with σ^2 covariance) to the Indian Pines and Pavia University dataset and show the impact of the noise to the proposed LGDRSR. We set TR = 10 per class and dim = 10. As can be seen from Fig. 9, the performance of the proposed LGDRSR decreases slightly when the noise level σ increases in Indian Pines. The proposed LGDRSR still performs very

robust in Pavia University dataset when the noise intensity is increased.

In addition, we conduct some experiments for comparing the proposed LGDRSR with some other state-of-the-art algorithms (we have used "-" in the Tables VII, VIII, and IX to denote the unavailable values in the references.), including spatial-spectral manifold reconstruction preserving embedding with k-nearest neighbors (SSMRPE-KNN) [44], the ensemble discriminative local metric learning with KNN and SVM (EDLML-KNN and EDLML-SVM) [45], fast dimensionality reduction and classification with extreme learning machine (FDRC-ELM) [46], and segmented stacked autoencoders (S-SAE) [47], respectively. The classification results for abovementioned state-of-the-art algorithms are directly taken from abovementioned references, respectively. As can be seen from Table VII, under the same

TABLE V
CLASSIFICATION ACCURACY (%) AND STANDARD DEVIATION FOR TESTING SAMPLES WITH 1% TRAINING SAMPLES PER CLASS FOR INDIAN PINES DATASET

NO.	TR	TS	DR+KSVM classifier								LGDRSR (m=0)	LGDRSR
			PCA	NWFE	LDE	RLDE	LFDA	SDA	SELDLPP	SELDNPE		
1	5	49	81.8±5.7	78.9±14.2	70.4±10.5	67.7±17.3	57.7±23.9	83.8±7.2	77.1±11.8	78.3±15.7	64.1±17.9	**98.9±1.4**
2	14	1420	60.1±8.7	57.4±9.1	56.0±10.2	50.4±7.8	38.5±10.8	55.0±11.0	49.9±9.9	55.9±7.0	54.8±3.9	**75.9±8.6**
3	8	826	48.3±6.3	54.0±5.9	39.9±8.4	35.3±7.9	31.3±11.0	46.9±8.5	43.6±8.3	49.8±6.9	41.0±8.9	**70.4±14.6**
4	5	229	43.0±12.3	53.7±12.4	24.6±16.4	33.4±12.4	34.4±12.9	45.2±15.6	45.7±14.5	50.3±14.9	41.7±10.0	**87.8±9.3**
5	5	492	62.5±9.9	57.0±14.4	59.5±8.9	58.6±10.7	61.9±7.0	68.0±7.9	63.5±8.7	65.1±7.2	64.5±16.0	**76.8±11.0**
6	8	739	75.6±7.0	80.5±6.0	74.2±13.7	57.8±10.6	65.3±5.8	78.1±8.9	75.9±8.1	76.0±9.4	78.1±7.3	**81.1±8.4**
7	5	21	89.5±5.8	90.9±11.0	92.3±9.3	78.5±19.0	81.9±13.2	90.9±5.7	90.9±2.7	90.4±7.0	86.5±3.8	**97.1±3.3**
8	5	484	69.1±11.8	65.8±11.1	73.4±13.0	62.4±14.9	52.2±17.6	69.4±13.0	62.9±13.1	68.2±13.7	86.9±16.2	**99.7±0.2**
9	5	15	90.0±14.1	97.3±3.4	72.6±26.1	62.0±14.0	82.0±14.0	**100±0**	91.3±13.3	94.0±11.5	88.8±13.0	**100±0**
10	10	958	60.8±10.6	59.0±15.8	47.2±18.0	43.4±9.4	37.7±5.9	50.8±14.6	51.3±15.2	58.3±15.4	55.0±8.6	**82.0±6.3**
11	24	2444	70.1±6.3	68.0±7.3	57.5±9.4	50.8±4.6	35.3±9.0	67.1±7.1	67.9±6.6	69.6±5.5	73.5±4.5	**90.6±2.7**
12	7	607	27.5±9.2	42.4±18.7	36.9±12.5	28.5±12.2	36.9±11.3	38.6±7.6	25.8±10.9	25.3±9.7	37.2±10.1	**73.8±9.8**
13	5	207	88.6±7.5	96.8±2.1	96.4±3.6	83.9±13.9	81.6±11.6	96.2±4.3	83.2±13.1	92.3±4.6	96.3±3.1	**99.5±0.2**
14	13	1281	85.7±6.4	86.1±9.1	86.5±6.9	83.6±7.6	78.1±10.3	82.2±13.4	85.2±8.1	86.0±7.4	88.6±8.3	**99.8±0.4**
15	5	375	24.2±5.2	34.5±7.8	33.4±10.6	34.5±7.7	38.7±10.8	36.5±13.3	21.1±6.1	22.1±10.1	27.4±8.6	**86.8±11.0**
16	5	95	86.7±4.0	87.0±5.7	87.4±6.1	85.1±7.1	84.1±4.4	87.6±4.4	87.7±4.6	87.6±4.1	89.5±4.5	**97.0±2.9**
OA			63.9±1.2	64.9±2.4	59.0±5.4	53.2±3.2	47.1±3.2	62.7±3.4	60.1±1.9	63.2±1.7	65.1±1.2	**85.5±1.8**
AA			66.5±1.7	69.3±2.4	63.0±5.3	57.2±2.9	56.1±2.0	68.5±2.3	63.9±1.1	66.8±2.0	67.2±2.5	**88.6±1.4**
k			58.7±1.4	59.9±2.9	53.3±6.0	46.9±3.6	40.9±3.3	57.5±3.7	54.3±2.1	57.9±2.0	60.0±1.4	**83.5±2.0**
DR (second)			0.02	1.21	0.04	0.11	0.04	0.83	0.76	1.48	7.10	8.96
Tr (second)			1.62	1.21	1.3	1.05	0.82	1.42	1.42	1.31		
Ts (second)			0.25	0.09	0.08	0.28	0.08	0.08	0.08	0.09	0.34	0.35

TABLE VI
CLASSIFICATION ACCURACY (%) AND STANDARD DEVIATION FOR TESTING SAMPLES WITH 1% TRAINING SAMPLES PER CLASS FOR PAVIA UNIVERSITY DATASET

NO.	TR	TS	DR+KSVM classifier								LGDRSR (m=0)	LGDRSR
			PCA	NWFE	LDE	RLDE	LFDA	SDA	SELDLPP	SELDNPE		
1	66	6565	84.0±1.9	86.5±1.8	84.4±1.5	86.7±2.0	85.4±2.6	87.5±1.9	85.8±1.2	84.7±1.8	81.4±2.0	**90.5±3.1**
2	186	18463	95.3±2.5	94.6±3.7	95.3±2.3	96.3±0.8	92.1±1.4	94.3±0.9	94.2±3.7	93.0±2.4	94.6±1.3	**99.7±0.1**
3	20	2079	44.1±13.2	57.8±8.9	45.0±13.9	58.8±10.6	51.7±5.5	60.9±7.9	48.7±9.6	49.2±12.1	47.9±5.5	**98.9±1.3**
4	30	3034	82.7±3.0	83.4±3.0	82.0±4.2	85.5±4.2	82.6±7.7	83.8±5.0	84.0±2.7	80.5±2.2	78.6±3.2	**94.1±1.2**
5	13	1332	98.9±0.5	99.2±0.4	98.3±2.7	97.3±6.0	99.4±0.2	99.4±0.3	99.4±0.1	97.7±2.1	99.3±0.1	**99.6±0.1**
6	50	4979	38.7±12.4	42.8±12.0	39.4±12.1	73.1±6.4	60.6±8.0	64.4±5.6	44.9±11.8	51.1±10.6	44.2±4.3	**99.8±0.1**
7	13	1317	70.8±8.7	76.1±8.5	69.7±9.8	73.1±10.0	59.4±11.2	73.2±8.6	70.5±8.0	74.2±6.1	70.1±6.9	**99.2±0.3**
8	37	3645	83.6±3.9	82.5±3.6	84.6±4.5	83.2±5.5	71.7±5.5	84.6±3.9	80.5±6.5	81.6±4.2	78.6±4.0	**98.1±0.7**
9	10	937	99.8±0.1	**99.9±0.1**	99.8±0.1	99.7±0.3	**99.9±0.1**	**99.9±0.1**	99.8±0.1	99.4±0.5	98.5±0.5	75.8±9.1
OA			81.9±0.9	83.3±0.9	82.1±1.0	87.7±1.1	82.3±1.4	86.2±0.6	82.5±0.9	82.5±0.9	81.3±0.6	**97.2±0.6**
AA			77.6±1.8	80.3±1.5	77.6±2.3	83.7±1.9	78.1±1.6	83.1±1.3	78.6±1.7	79.0±2.0	77.0±1.0	**95.1±1.2**
k			75.4±1.5	77.3±1.3	75.6±1.6	83.5±1.6	76.3±1.9	81.4±0.9	76.3±1.2	76.4±1.3	74.6±0.9	**96.3±0.8**
DR (second)			0.03	1.47	0.17	3.53	0.06	0.59	0.47	0.88	55.57	74.82
Tr (second)			16.03	13.09	15.27	3.56	5.71	5.39	10.03	9.19		
Ts (second)			0.63	0.61	0.66	0.62	0.58	0.59	0.64	0.87	0.5	0.60

situations, the proposed LGDRSR has obtained a better performance than these methods. Also, we have conducted experiment in Salinas dataset [48]. The size of this dataset is 512×217 with a ground resolution of 3.7 m, and the number of the band and the land cover types are 204 and 16, respectively. We choose about 1.5% samples for training and remaining for testing. As can be seen from Table VIII, the proposed LGDRSR has achieved better classification accuracies than scale-orientation morphological profiles (SOMPs) [49] based

methods (the classification results are also directly taken from reference [49] as well), such as SOMP-spectral angle distance (SOMP-SAD), SOMP-spectral information divergence (SOMP-SOD), and SOMP-hidden Markov model-based information divergence (SOMP-HMMID). This verifies the validity of the proposed LGDRSR.

Furthermore, the Table IX reported some additional experiments to verify the proposed DRSR and LGDRSR, as can be seen from Table IX, under the setting with five samples per class and
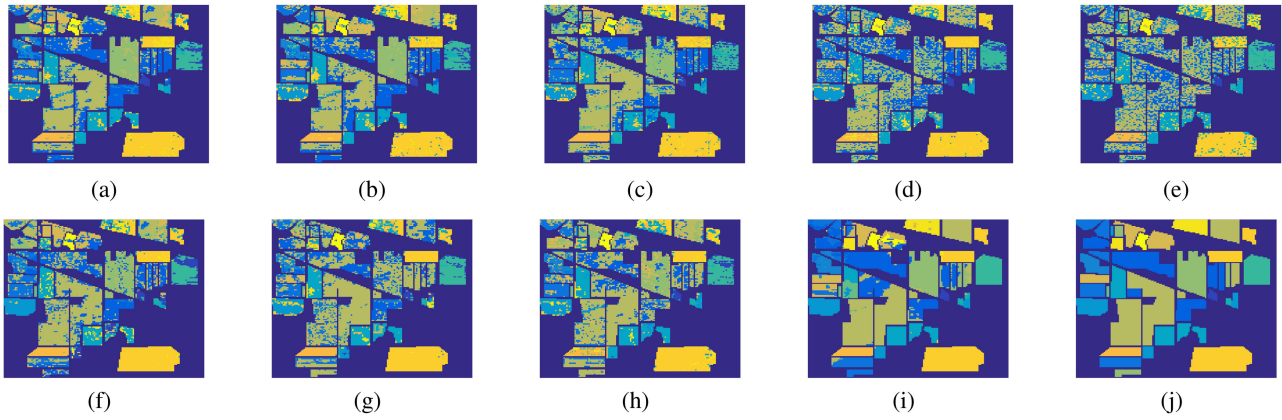
Fig. 7.　Classification accuracies on Indian Pines dataset. (a) KSVM-PCA. (b) KSVM-NWFE. (c) KSVM-LDE. (d) KSVM-RLDE. (e) KSVM-LFDA. (f) KSVM-SDA. (g) KSVM-SELDLPP. (h) KSVM-SELDNPE. (i) LGDRSR. (j) Ground Truth.
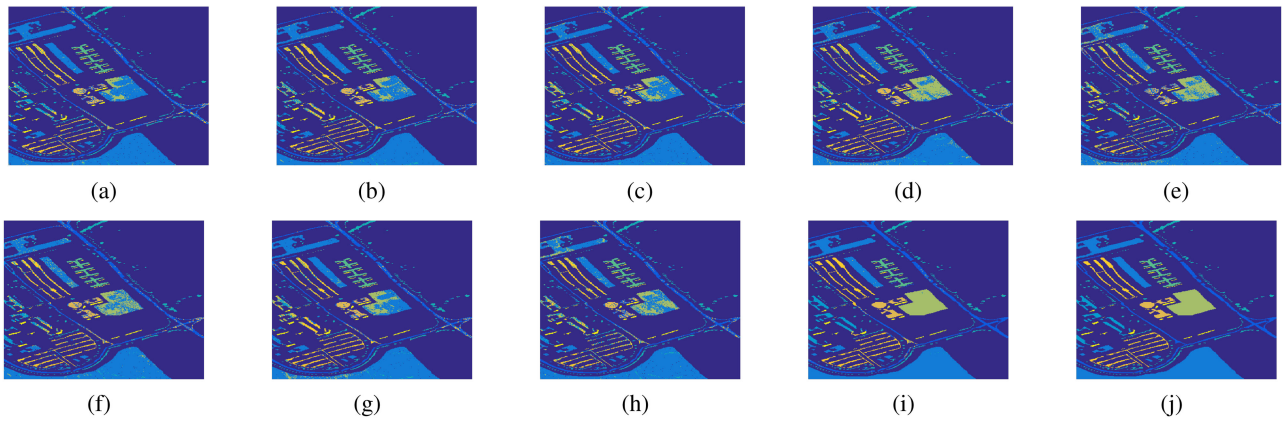

Fig. 8.　Classification accuracies on Pavia University dataset. (a) KSVM-PCA. (b) KSVM-NWFE. (c) KSVM-LDE. (d) KSVM-RLDE. (e) KSVM-LFDA. (f) KSVM-SDA. (g) KSVM-SELDLPP. (h) KSVM-SELDNPE. (i) LGDRSR. (j) Ground Truth.

TABLE VII
CLASSIFICATION ACCURACY (%) AND STANDARD DEVIATION FOR INDIAN PINE AND PAVIA UNIVERSITY DATASET

Dataset	Index	TR number: 20 per class; Dim: 30		TR number: 10% for training; Dim: 61			TR number: 1% for training; Dim: 40		TR number: 5% for training; Dim: 20	
		SSMRPE-KNN	LGDRSR	EDLML-KNN	EDLML-SVM	LGDRSR	FDRC-ELM	LGDRSR	S-SAE	LGDRSR
Indian Pines	OA	-	90.99±0.81	76.3±0.48	85.86±0.69	97.1±0.4	64.91±1.17	85.5±1.8	80.66	96.0±0.6
	AA	-	94.58±0.69	-	-	94.6±2.0	-	88.6±1.4	73.63	95.1±1.2
	k	-	89.76±0.91	72.59±0.54	83.86±0.73	96.7±0.4	-	83.6±2.0	-	95.4±0.7
Pavia University	OA	86.92±1.22	90.5±1.9	90.70±0.14	96.57±0.75	99.1±0.1	82.83±0.05	96.1±0.6	-	98.82±0.16
	AA	-	92.0±0.8	-	-	98.4±0.2	-	92.4±1.3	-	97.52±0.34
	k	-	87.6±2.4	87.52±0.19	94.78±0.81	98.8±0.1	-	94.7±0.8	-	98.43±0.22

TABLE VIII
CLASSIFICATION ACCURACY (%) AND STANDARD DEVIATION FOR SALINAS DATASET

Dataset	Index	1.5% training samples, remaining for testing				
		Dim:13		Dim:14		
		SOMP-SAD	LGDRSR	S0MP-SID	SOMP-HMMID	LGDRSR
Salinas	OA	94.34	98.82±0.20	95.27	95.03	98.81±0.20
	AA	92.81	98.24±0.28	94.24	93.65	98.23±0.27
	k	-	98.69±0.22	-	-	98.67±0.22

TABLE IX
CLASSIFICATION ACCURACY (%) AND STANDARD DEVIATION FOR INDIAN PINE AND PAVIA UNIVERSITY DATASET

| Dataset	Index	5 samples per class; Dim:5		Dim:15; 20% training set for FA-MEC and FA-JM; 1% training set for LGDRSR; remaining for test		
		DRSR-without TS	DRSR	FA-MEAC	FA-JM	LGDRSR
Indian Pines	OA	48.12±3.82	50.41±4.60	-	-	85.99±1.45
	AA	52.46±1.76	53.12±2.63	-	-	88.10±1.77
	k	41.88±3.69	44.60±4.48	-	-	84.06±1.67
Pavia University	OA	51.66±5.38	54.61±8.15	91.91±0.01	91.95±0.00	96.23±0.60
	AA	47.84±5.51	50.87±5.85	-	-	92.89±1.33
	k	39.41±5.30	42.26±8.23	89.2±0.01	89.3±0.01	95.00±0.80
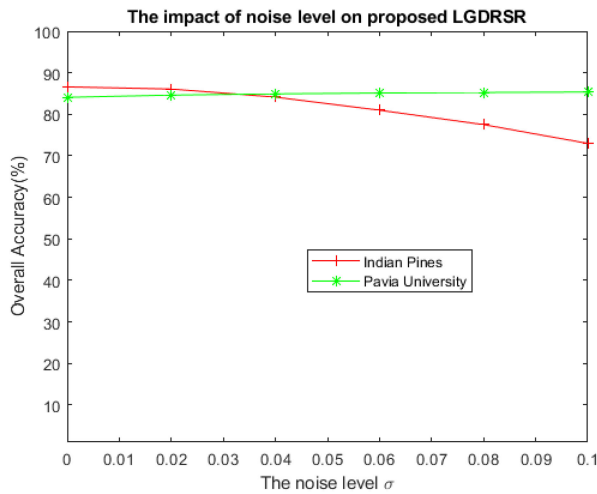
Fig. 9. Impact of noise level on proposed LGDRSR.

five dimensions, we can seen the proposed DRSR has achieved better classification accuracies than DRSR without testing samples (USDR). This verified the proposed DRSR; and the Table IV also shows the classification accuracies of well-known firefly algorithms (FA) based band selection method [50], such as, FA-minimum estimated abundance covariance (FA-MEAC) and FA with Jeffreys–Matusita distance (FA–JM). As can be seen from Table IX, with the same dimensions (15), the proposed LGDRSR has obtained better results than these methods even the training samples are less than FA-MEAC and FA-JM. Hence, these verified the proposed DRSR and LGDRSR.

V. CONCLUSION

In this article, a novel framework based on local and global dimensionality reduction for SRC has been proposed to simultaneously reduce the dimensions of HSIs and classify them. By constructing the DRSR method, which extracts the most discriminative features from data and classifies the classes HSI images, the proposed DRSR has demonstrated improvement on the overall classification performance. Furthermore, the local and global information of HSIs to integrate the spectral and spatial information of HSIs has greatly improved the classification accuracies. The experiments results have shown the outstanding performance our proposed method achieved compared to other state-of-the-art methods. In comparison with some DR+ SVM methods with 1% training samples, the improvement is more than 20% and 10% in Indian Pines and Pavia University dataset, respectively.

For the future work, we will investigate the second-order Taylor series [51] to further improve the recognition rate of the proposed methods. Besides, some mathematical methods, such as singular value decomposition, algebraic property, and probabilistic method [52], will be applied to the proposed methods for improving the computational efficiency.

REFERENCES

[1] H. Huang and M. Yang, "Dimensionality reduction of hyperspectral images with sparse discriminant embedding," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5160–5169, Sep. 2015.

[2] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-4, no. 1, pp. 55–63, Jan. 1968.

[3] B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson, "Band selection for hyperspectral image classification using mutual information," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 522–526, Oct. 2006.

[4] F. Cao *et al.* "Extreme sparse multinomial logistic regression: A fast and robust framework for hyperspectral image classification," *Remote Sens.*, vol. 9, no. 12, 2017, Art. no. 1255.

[5] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.

[6] J. Wright, A. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[7] F. Cao, Z. Yang, J. Ren, W. Chen, G. Han, and Y. Shen, "Local block multilayer sparse extreme learning machine for effective feature extraction and classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5580–5594, Aug. 2019.

[8] J. Liu, Z. Wu, J. Li, A. Plaza, and Y. Yuan, "Probabilistic-kernel collaborative representation for spatial-spectral hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2371–2384, Apr. 2016.

[9] P. Zhou, Z. Lin, and C. Zhang, "Integrated low-rank-based discriminative feature learning for recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 5, pp. 1080–1093, May 2016.

[10] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[11] Y. Li, Z. L. Yu, N. Bi, Y. Xu, Z. Gu, and S. Amari, "Sparse representation for brain signal processing: A tutorial on methods and applications," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 96–106, May 2014.

[12] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[13] Y. Chen, M. Nasser, and D. Trac, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[14] Y. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 296–308, Jan. 2015.

[15] Y. Chen, M. Nasser, and D. Trac, "Sparse representation for target detection in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 629–640, Jun. 2011.

[16] H. Huang and M. Yang, "Dimensionality reduction of hyperspectral images with sparse discriminant embedding," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5160–5169, Sep. 2015.

[17] P. Thouvenin, N. Dobigeon, and J. Tourneret, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 525–538, Jan. 2016.

[18] S. Chen and D. Zhang, "Semisupervised dimensionality reduction with pairwise constraints for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 369–373, Mar. 2011.

[19] X. He and N. Partha, "Locality preserving projections," *Adv. Neural Inf. Process. Syst.*, vol. 16, no. 16, pp. 153–160, 2004.

[20] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, 2005, vol. 1, no. 2, pp. 1208–1213.

[21] J. Zabalza *et al.* "Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing," *ISPRS-J. Photogramm. Remote Sens.*, vol. 93, pp. 112–122, 2014.

[22] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Mullers, "Fisher discriminant analysis with kernels," *Neural Netw. Signal Process. IX: Proc. IEEE Signal Process. Soc. Workshop*, 1999, pp. 41–48.

[23] B. Kuo and D. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 5, pp. 1096–1105, May 2004.

[24] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, pp. 1027–1061, 2007.

[25] B. Raducanu and F. Dornaika, "A supervised non-linear dimensionality reduction approach for manifold learning," *Pattern Recognit.*, vol. 45, no. 6, pp. 2432–2444, 2012.

[26] H. Chen, H. Chang, and T. Liu, "Local discriminant embedding and its variants," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 2, pp. 846–853.

[27] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, 2007, pp. 1–7.

[28] A. Lagrange, M. Fauvel, S. May, and N. Dobigeon, "Matrix cofactorization for joint spatial-spectral unmixing of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4915–4927, Jul. 2020.

[29] L. Zhang, M. Yang, Z. Feng, and D. Zhang, "On the dimensionality reduction for sparse representation based face recognition," in *Proc. 20th Int. Conf. Pattern Recognit.*, 2010, pp. 1237–1240.

[30] S. Kim *et al.* "A method for large-scale l1-regularized least squares," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 606–617, Dec. 2007.

[31] Q. Wang, X. He, and X. Li, "Locality and structure regularized low rank representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 911–923, Feb. 2019.

[32] DP. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. New York, NY, USA: Academic, 2014.

[33] K. Ito and K. Karl, "An active set strategy based on the augmented lagrangian formulation for image restoration," *ESAIM: Math. Modelling Numer. Anal.*, vol. 33, no.1, pp. 1–21, 1999.

[34] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.

[35] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947–3960, Oct. 2011.

[36] X. Xu, J. Li, S. Li, and A. Plaza, "Subpixel component analysis for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5564–5579, Aug. 2019.

[37] L. Fang, G. Liu, S. Li, P. Ghamisi, and J. Benediktsson, "Hyperspectral image classification with squeeze multibias network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1291–1301, Mar. 2019.

[38] Q. Hao, S. Li, and X. Kang, "Multilabel sample augmentation-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4263–4278, Jun. 2020.

[39] S. Alim, J. Li, C. Lin, S. Liu, and E. Li, "Edge gradient-based active learning for hyperspectral image classification, " *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 9, pp. 1588–1592, Sep. 2020.

[40] Y. Zhou, J. Peng, and C. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1082–1095, Feb. 2015.

[41] W. Liao, A. Pizurica, P. Scheunders, W. Philips, and Y. Pi, "Semisupervised local discriminant analysis for feature extraction in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 184–198, Jan. 2013.

[42] C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2011, Art. no. 27.

[43] [Online]. Available: https://www.fst.um.edu.mo/en/staff/fstycz.html

[44] H. Huang, G. Shi, H. He, Y. Duan, and F. Luo, "Dimensionality reduction of hyperspectral imagery based on spatial-spectral manifold learning," *IEEE Trans. Cybern.*, vol. 50, no. 6, pp. 2604–2616, Jun. 2020.

[45] Y. Dong, B. Du, L. Zhang, and L. Zhang, "Dimensionality reduction and classification of hyperspectral images using ensemble discriminative local metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2509–2524, May 2017.

[46] J. M. Haut *et al.*, "Fast dimensionality reduction and classification of hyperspectral images with extreme learning machines," *J. Real-Time Image Process.*, vol. 15, no. 3, pp. 439–462, 2018.

[47] J. Zabalza *et al.*, "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging," *Neurocomputing*, vol. 185, pp. 1–10, 2016.

[48] F. Hou *et al.*, "FMRSS net: Fast matrix representation-based spectral-spatial feature learning convolutional neural network for hyperspectral image classification," *Math. Problems Eng.*, vol. 2008, 2018, Art. no. 9218092.

[49] A. Plaza *et al.*, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466–479, Mar. 2005.

[50] H. Su, B. Yong, and Q. Du, "Hyperspectral band selection using improved firefly algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 68–72, Jan. 2016.

[51] F. Cao *et al.*, "Sparse representation-based augmented multinomial logistic extreme learning machine with weighted composite features for spectral-spatial classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6263–6279, Nov. 2018.

[52] Z. Yang, F. Cao, Y. Cheng, W.-K. Ling, and R. Hu, "Locality regularized Robust-PCRC: A novel simultaneous feature extraction and classification framework for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8567–8582, Dec. 2020.

Faxian Cao (Student Member, IEEE) was born in Jiangxi, China, in September 1993. He received the B.E. degree in electronics and information engineering from Jinggangshan University, Ji.an, China, in 2016, and the M.E. degree in electronics communication engineering from the Guangdong University of Technology, Guangzhou, China, in 2019. He is currently working toward the Ph.D. degree in electronics engineering with the School of Electronic Engineering and Computer Science, Queen Mary, University of London, London, UK.

He was a Visiting Researcher with the University of Strathclyde, Glasgow, Scotland, from November 2018 to May 2019. His main research interests comprise signal processing, ad hoc time delay estimation, echo SLAM, machine learning, and pattern recognition and image processing.

Mr. Cao is currently a Reviewer of the IEEE GEOSCIENCE AND REMOTE SENSING LETTER, IEEE ACCESS, *International Journal of Remote Sensing*, and *European Journal of Remote Sensing*, etc.

Zhijing Yang (Member, IEEE) received the B.S. and Ph.D. degrees in mathematics and computing science from Sun Yat-Sen University, Guangzhou, China, in 2003 and 2008, respectively.

He was a Visiting Research Scholar with the School of Computing, Informatics and Media, University of Bradford, Bradford, U.K., between July–December 2009, and a Research Fellow with the School of Engineering, University of Lincoln, Lincoln, U.K., between 2011– 2013. He is currently a Professor and the Deputy Dean with the School of Information Engineering, Guangdong University of Technology, Guangzhou, China. His current research interests include image retrieval, machine learning, and pattern recognition.

Yuezhen Huang was born in September 1973 in Zhangpu, Fujian. He received the master's degree in business administration from the South China University of Technology, Guangzhou, China.

He is a CPC member with Han nationality. He is currently the President of Guangzhou Radio Group Company, Ltd., Guangzhou, China, the Chairman of GRGBanking, Guangzhou, China, and the Chairman of the Group Science and Technology Association.

Xiaobin Hong received the Ph.D. degree in mechanical engineering from the South China University of Technology, Guangzhou, China, in 2007.

He is currently a Professor with the School of Mechanical and Automotive Engineering, South China University of Technology. His research interests include instrumentation design, signal processing, and machine learning methods for nondestructive detection and structural health monitoring.

Jujian Lv received the Ph.D. degree in computer application technology from Sun Yat-Sen University, Guangzhou, China, in 2016.

He is currently a Lecturer with the School of the Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China. His research interests include image processing, pattern recognition, and machine learning.
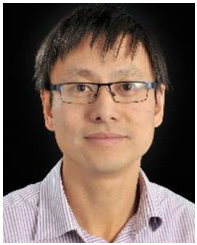
Yongqiang Chen received the B.Eng. and M.Eng. degrees in control theory and control engineering from Tongji University, Shanghai, China, in 2001 and 2004, respectively, and the Ph.D. degree from the School of Engineering, Design and Technology, University of Bradford, U.K., in 2010.

He is currently a Senior Lecturer with the Department of Computer Science and Technology, University of Hull, Hull, U.K. Before this, he was a Postdoctoral Research Fellow with the Future Ubiquitous Networking Laboratory, School of Engineering and Informatics, University of Bradford, Bradford, U.K., from 2010 to 2014. His research includes digital healthcare technologies, AI and machine learning, embedded system, control theory, and applications.